Step 2: Running Extract Validation
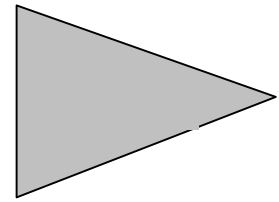
Once you have created your Extract file, you're ready for step 2, running Extract Validation. This step is done entirely by DataPrep.

What Happens in Extract Validation?

In the Extract Validation process, DataPrep first examines the Extract file to make certain the format is acceptable. DataPrep checks for a proper header record, 300-byte record lengths, and school code. These are called file-level edits. If the header format is not correct, DataPrep cannot continue the process and an error message will appear in a dialog box informing you that there was a header error and that processing was aborted. The Extract Validation process will also abort if any record has a school code that does not match the header record school code. (Therefore, servicers and schools reporting multiple schools or branches should be certain there is a header for each school or branch.)

If your Database Extract file passes the file-level edits, DataPrep examines all Detail and Past Period Change records in the file to ensure that each data element meets domain requirements. If the percent of domain errors exceeds the threshold set by ED (see box at right), DataPrep will issue an error message informing you that you have exceeded the threshold and that no Submittal file was created. All errors are noted in an Extract Error Data file from which you can generate a report detailing the error records. From this you can correct your database or extract program and create a new Extract file. You then rerun the Validation Extract program with your new Extract file to create the Submittal file.

If your Database Extract file passes the file-level edits and the percent of domain errors is below the maximum threshold established by NSLDS, DataPrep will create a Submittal file that you will then send to NSLDS.

Reasons Extract Validation Will Abort
1. File-level errors
2. Header record errors

(See Appendix B-10 (PC users) or B-11 (OS 390/LE users) for a description of these types of errors.)

Domain-Level Errors
There are four kinds of domain-level errors:
1. Numeric field errors (character other than a number is in a field requiring all numbers)
2. Invalid date errors (date specified does not exist on a calendar or is not zeros)
3. Missing Identifiers
4. Missing New Identifiers

The Extract Validation process produces three output files:

1. **Validation Log File**—a summary report of the transactions processed during Extract Validation,

2. **Extract Error File**—a file from which you can generate a report listing all domain errors (created only if the Database Extract file passes file-level edits), and

3. **Submittal File**—the file you will transmit to NSLDS (created only if the Database Extract file passes file-level and domain-level edits).
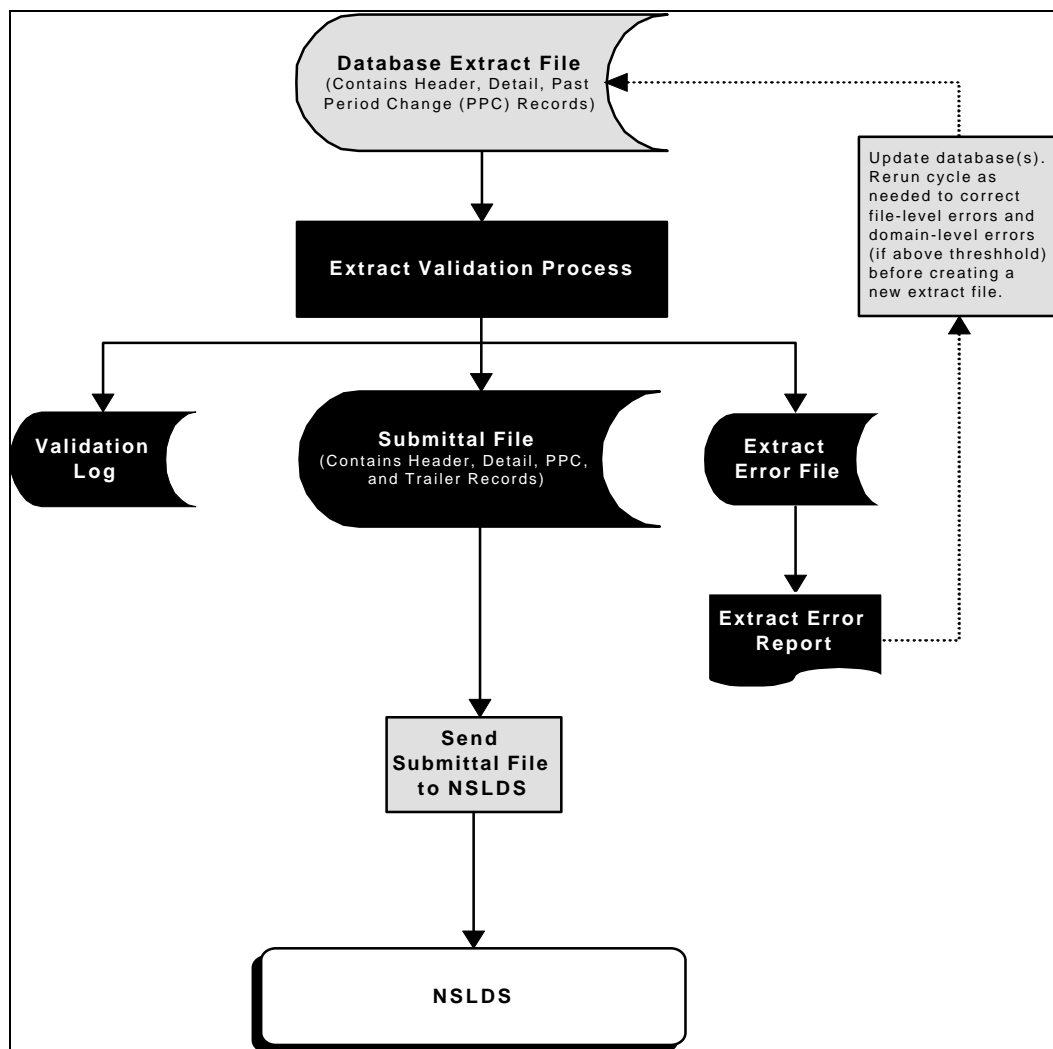

Figure 1, Extract Validation Process

DataPrep Error Path

DataPrep performs two sets of edits during the Extract Validation process:

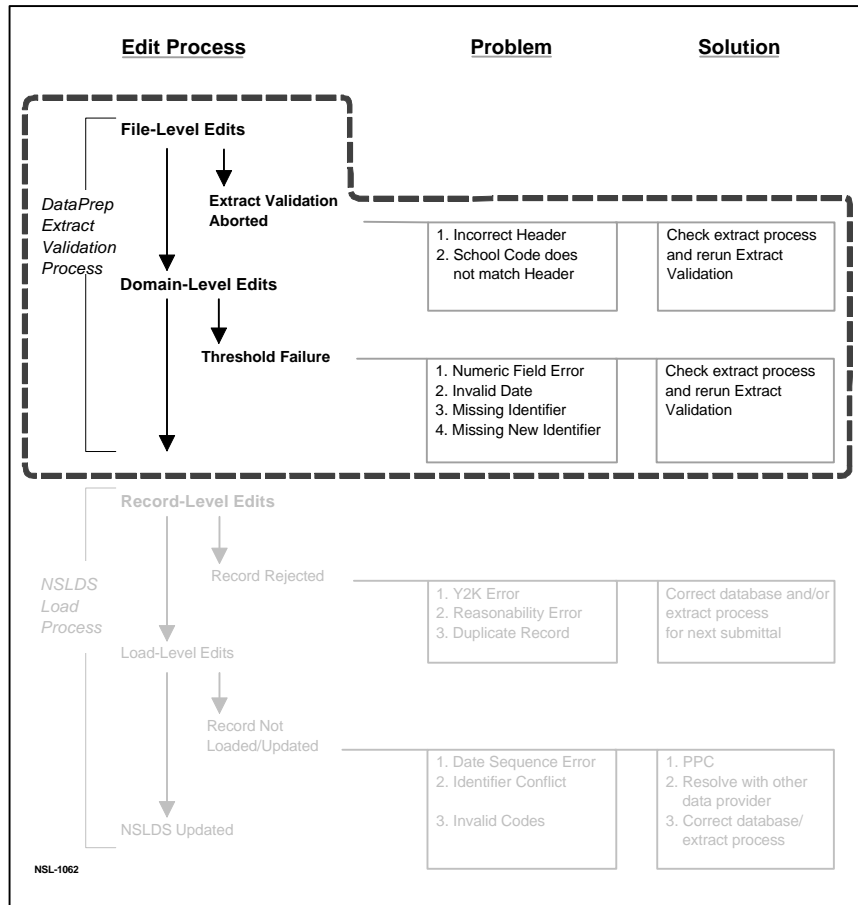1. File-level edits
2. Domain-level edits

Figure 2, DataPrep Edit Process

File-Level Edits

File-level edits check to see that the Extract file is a legitimate file with the correct header, 300-byte records, and a school code in each record that matches the header. If DataPrep detects any one of these file-level edits, the Extract Validation process will abort and an error message, with a description of the error, will appear on screen. You must correct your Extract file and rerun Extract Validation. See Appendix B-10 (PC users) or B-11 (OS 390/LE users) for a complete list of all the file-level and header errors that cause the Extract Validation process to abort.

Domain-Level Edits

There are four kinds of domain-level errors: numeric field errors (character other than a number is in a field requiring all numbers), invalid date errors (date specified does not exist on a calendar or is not all zeros), missing identifiers, and missing new identifiers. If the percent of these errors exceeds the threshold established by NSLDS, the Extract Validation process will not be completed and no Submittal file will be created. You must then correct your database and rerun the Extract Validation process.

Multiple Errors in a Record
DataPrep validates the entire record and multiple errors can be detected on a single input record. The error rate calculated by DataPrep is based on the number of records with one or more errors, not on the total number of errors detected.

Running Extract Validation on a PC

Once you have installed the DataPrep software and set up all the appropriate files and directories, you are ready to perform Extract Validation.

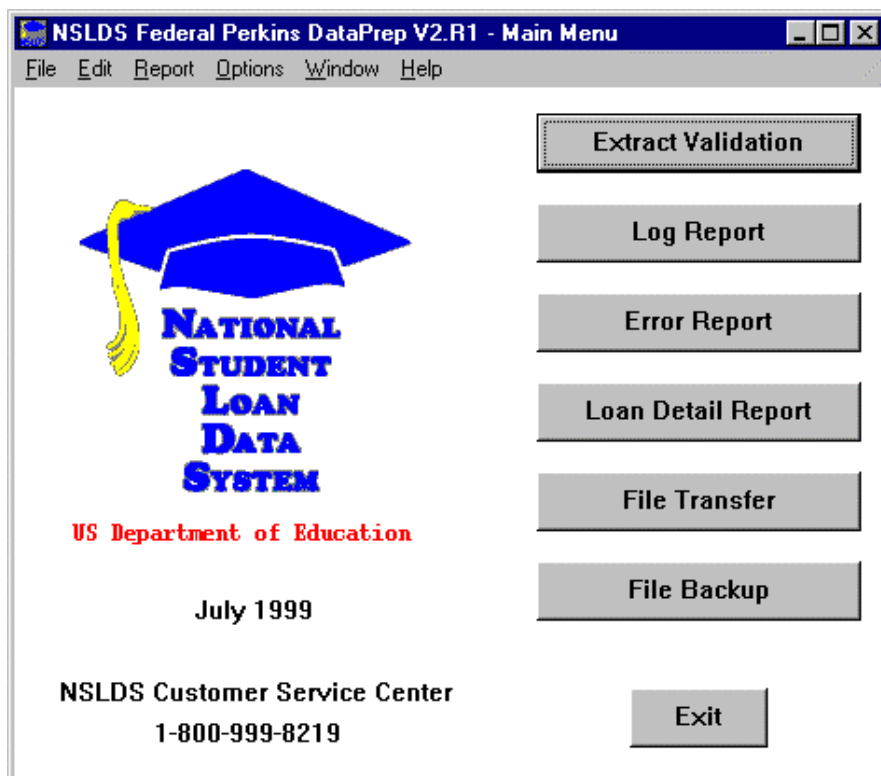To begin, on the DataPrep Main Menu, click Extract Validation.

Figure 3, DataPrep Main Menu with Extract Validation Selected

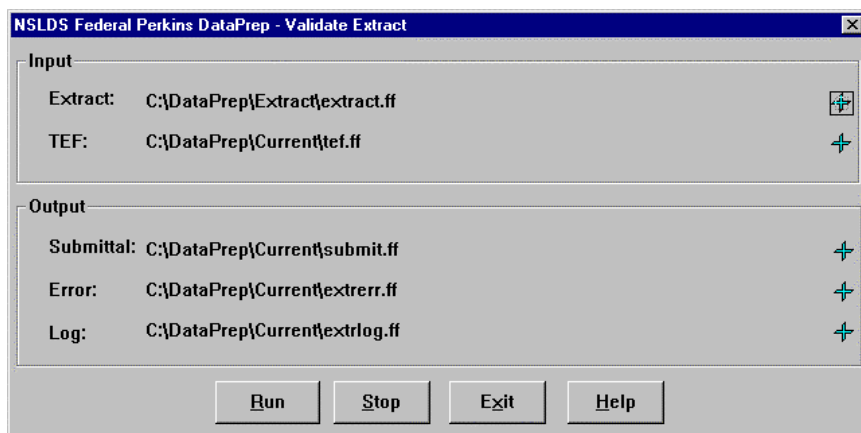The Validate Extract dialog box will appear.

Figure 4, Validate Extract Dialog Box

⚠️ **Naming the Extract File**
Remember that your database Extract file must be named extract.ff so that DataPrep can locate and process it.

💡 **File Date**
Note that the date a file was last modified or created appears on the screen. This is to help you make sure you are running the correct Extract file.

If you select the plus sign next to the file date, the File Information Dialog Box will appear showing you the date and time the file was last modified and the number of bytes in the file.

If you select any of the plus signs on the right of the screen next to the file date, the File Information dialog box will appear. The box shows the file name, the date and time the file was created or last modified—whichever is more current—and the number of bytes in the file.
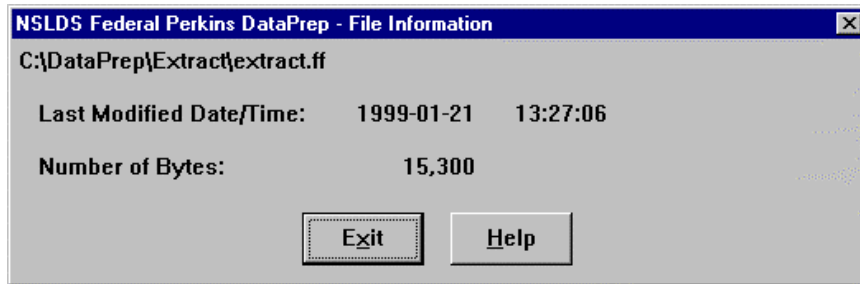

Figure 5, File Information Dialog Box

Extract Validation

Since you have already defined the location of your Database Extract and Threshold, Error Code, and Field Code (TEF) files—see How to Install DataPrep—and DataPrep has defined for you the output files, you are now ready to Run Extract Validation. When you do, DataPrep will perform Extract Validation on your Database Extract file and will show the progress. You must be certain that your Extract file was put in the correct directory (e.g., C:\DataPrep\Current\extract.ff) and was named correctly (it must be named extract.ff) so that DataPrep can locate the file.

Once Extract Validation begins, the Validate Extract Status screen will appear (see Figure 6 below), showing you how much of the process is complete. While the process is being performed, you may go on to other tasks by minimizing the screen or by selecting Close. When the process is complete, the Validate Extract Status screen will show that 100% of the process was complete, and will detail the number of Extract Detail records processed, the number of Past Period Change records, Date/Numeric Errors, Identifier Errors, and New Identifier Errors.

Once you close this screen, you'll get back to the previous screen from which you must exit to get back to the DataPrep Main Menu. From there you have several options, including generating reports.

Number of Bytes,
***Not* Number of Records**
The File Information Dialog Box shows the number of bytes in the file, *not* the number of records. For the Extract file, dividing by 300 bytes will yield the number of records. But for any error or report file, you cannot easily determine the number of records in the file.

While Validation is Processing
While DataPrep is performing the Validation process, you may minimize the screen and go on to other tasks. When the process is complete, the Validate Extract Status screen will show that 100% of the process was completed.
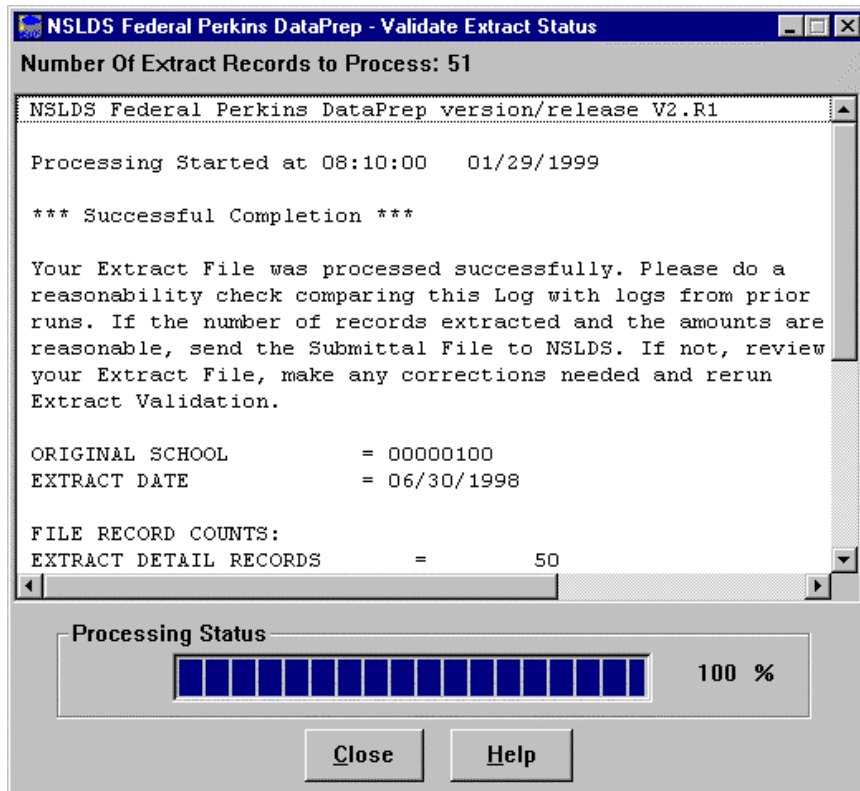
Figure 6, Validation Extract Processing Status

Output

The successful Extract Validation process produces three files:

1. Validation Log file (extrlog.ff),
2. Extract Error file (extrerr.ff), and
3. Submittal file (submit.ff).

Validation Log

The Validation Log is a summary report of the transactions processed during Extract Validation. It also tells you the percentage of domain errors (combined date and numeric field errors, identifier errors, and new identifier errors) in your Extract file, as well as loan totals. The Validation Log can either be viewed on screen or printed through the viewer program software you designated in the setup procedures (under Options). You can also change the viewer designation by selecting Viewer.

To view a Validation Log, first go to the Main Menu and click Log Report.

Figure 7, DataPrep Main Menu with Log Report Selected
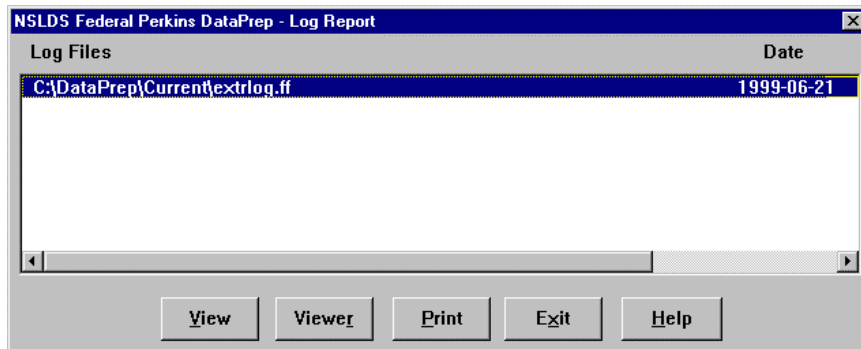
The Log Reports dialog box will appear:

Figure 8, Log Reports Dialog Box

From this screen you must first highlight the appropriate file (the log file in your current directory). After that, you can either print the report or view it using a viewer you designate.

Designated Viewer
When you first set up DataPrep, you can specify which viewer program software you'll use to both view and print the reports (either the DataPrep default software, MS-Wordpad, or Notepad). In addition, you can change which viewer you use by clicking on the Viewer button and changing the designation. This does not change the default viewer —the viewer you select will only be used for this one report.

```
extrlog.ff - WordPad                                    _ □ ×
File  Edit  View  Insert  Format  Help

 [icons]

 NSLDS Federal Perkins DataPrep version/release V2.R1

 Validation Processing Started at 11:55:04   06/21/1999

 *** Exceeds Threshold Tolerance ***
 The percentage of domain errors exceeds the allowable
 tolerances. Therefore, no Submittal File has been created.
 You may use the Loan Detail Error Report to help determine
 the cause. Please correct your database, create a new
 Extract file, and rerun the Extract Validation process.
 Refer to the Perkins DPI for help in identifying the
 possible cause of the problem.

 ORIGINAL SCHOOL        = 00111100
 EXTRACT DATE           = 06/30/1998

 FILE RECORD COUNTS:
    EXTRACT DETAIL RECORDS    =           50
    PAST PERIOD CHANGE RECORDS =           0

 DATE/NUMERIC ERRORS         =           1    2.00%
 IDENTIFIER ERRORS           =           5   10.00%
 **IDENTIFIER ERRORS EXCEED THE ALLOWABLE PERCENTAGE**

 NEW IDENTIFIER ERRORS       =           0    0.00%

 TOTALS FOR OPEN LOANS:
    NUMBER OF OPEN LOANS         =           20
    AMOUNT OF LOAN               =       31,922
    AMOUNT OF CANCELLATION       =            0
    OUTSTANDING PRINCIPAL BALANCE =      25,882

 Validation Processing  Ended  at 11:55:06   06/21/1999
 Total Processing Time:   00 Hours 00 Minutes 02 Seconds

For Help, press F1                                      NUM
```

Figure 9, Validation Log

Using the Information from the Validation Log File

The Validation Log file contains four types of information:

1. An explanation of your Extract Validation process, including whether the process succeeded or failed and instructions about what you should do next.

2. Extract File Summary Data,

3. The records processed, with an accounting of the domain-level errors, and

4. Summary loan data to provide a monthly reference.

Explanation of Your Extract Validation Process

This information tells you whether the process was successful and provides instructions about what you should do next.

⚠

Entire Submittal Rejected for High Error Rate
If the number of domain errors (date/numeric field error, identifier errors, or new Identifier errors) found in the Detail and PPC records exceeds the threshold as defined by ED, the *whole* Submittal file will be rejected! The Validation Log File provides you with your error rates.

Extract Validation is Successful

If your Extract file was processed successfully, the Log will state:

> Your Extract file was processed successfully. Please do a reasonability check comparing this Log with Logs from prior runs. If the amounts and number of records extracted are reasonable, send the Submittal file to NSLDS. If not, review your Extract file, make the needed corrections, and rerun Extract Validation.

Compare the summary data on this Log with the summary data on Logs you ran for prior Validation Extracts to make sure the numbers are reasonable. Look at the number of records processed, amount of loan, and amount of outstanding principal to make sure they are close to the numbers for prior Extracts.

Extract Validation Fails Because of File-Level Errors

If DataPrep was unable to read your Extract file or if the Validation Process was aborted because of other file-level errors, the Validation Status screen will automatically appear and state the reason the process was halted.
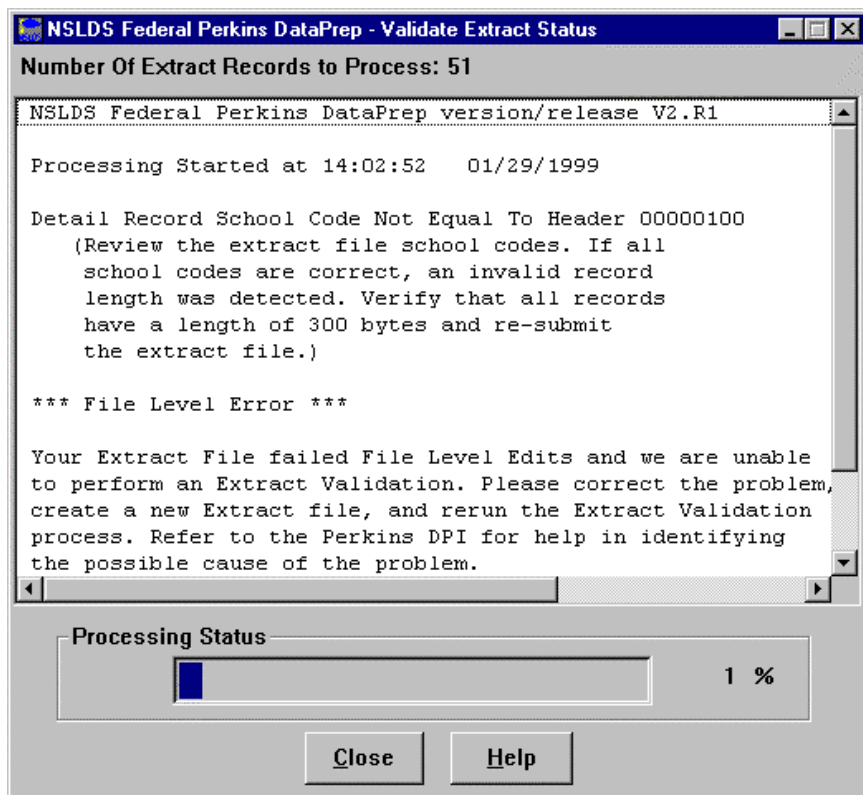

Figure 10, Failed Validate Extract Status Screen

Some possible causes of a failed Validation are:

- No header record

- An incorrect format was detected

- Your data shifted because you inserted a space or a character

- The records were not the required 300-byte length

- The code for original school does not match the school code in the header record.

Extract Validation Fails Because of Domain Errors

If the percentage of domain-level errors exceeded the allowable threshold, the Log will state that no Submittal file was created, will report the error rate, and will explain the reason for the failure (excessive date/numeric, identifier, or new identifier errors). The Validation Log will state:

> The percentage of domain errors exceeds the allowable tolerances. Therefore, no Submittal File has been created. You may use the Loan Detail Error Report to help determine the cause. Please correct your database, create a new Extract file, and rerun the Extract Validation process. Refer to the Perkins DPI for help in identifying the possible cause of the problem.

When you receive this message, you must correct the domain-level errors on your database so that the percentage of errors is acceptable. You can use the Loan Detail Error Report to see what corrections to your Database Extract file must be made.

There are a number of reasons for such errors. Some of the following causes and corrections may explain your situation:

- Your data is stored incorrectly on your database: the solution is to correct the appropriate fields on your database (i.e., your database accepted six digit dates instead of eight).

- Your extract process calculates fields incorrectly: review and correct any programming logic in your extract process (i.e., Date Entered Repayment is calculated by adding one day to the end of the enrollment period and sets the date to 2-29-1999, an invalid date, instead of 3-1-1999).

- Your extract process only picks up changed fields: change your process to populate the other fields with the current data for those fields.

Extract File Summary Data

The Extract File Summary Data contains the Extract Date and the OPEID of the school being processed. When multiple institutions are reported on the same extract the School ID field is populated by the number of schools reported on the file. The Extract Date is provided for your reference on the Log.

Records Processed

Records Processed provides key information for review with each submission, including the numbers of Detail records and Past Period Change records. It also contains the numbers of domain errors (Date/Numeric Errors, Identifier Errors, and New Identifier Errors) found in the Detail and Past Period Change records. If the percentage of any of these errors exceeds the threshold defined by ED, DataPrep will not create a Submittal file. The Log file will provide your error rate and will tell you if you've exceeded the acceptable threshold. If so, you must correct your database, re-start the extract process, and re-validate until your error rate is below the accepted level.

Loan Record Summary

The Loan Record Summary provides basic totals for your review. Totals for Amount of Loan, Amount of Cancellation, and Outstanding Principal Balance are provided to help you spot errors or potential problems. You should review this summary before transmitting your Submittal file to see if there are any glaring errors or large changes that could indicate a problem in the data or your extract process.

Checking the Loan Record Summary
Large increases in the number of loans, amounts of loans, or outstanding principal balance could indicate duplication of records or extracting some incorrect records.

Extract Validation for Mainframes (OS/390 LE-Based Users)

The JCL for Mainframes (OS/390 LE-Based Users) performs Extract Validation and error file generation. Appendix G contains the JCL for these functions. It can be referenced from the library created with JCLLIB as part of the name. The library member name is DRBB1000.

The JCL references a sample Extract file containing 50 student/loan records of which 2 are in error. This should be reported in the Log file report, the Detail Error Report, and the Summary Error Report that is output from the Extract Validation process.

A second sample extract is included for your use that contains 50 student/loan records, of which 6 are in error, and in which error thresholds are exceeded. If you use this sample to test your installation, no Submittal file should be created. To use the second sample, you must change the JCL to reference the sample extract containing DBEXTERR as part of the name.

⚠

Previous Datasets
The first step in the JCL will delete any datasets previously created. If you want to save your previous Submittal file, you must copy it to another file name.